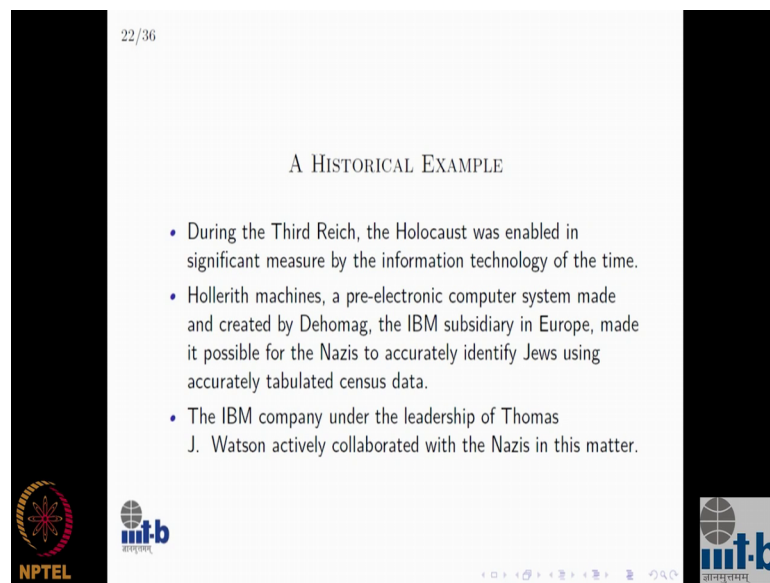


**Digital And The Everyday: From Codes To Cloud**  
**Prof. Shrisha Rao**  
**Department of Multidisciplinary**  
**International Institute of Information Technology, Bangalore**

**Lecture - 05**  
**Socio-algorithmic processes & the Everyday-Part 03**

This is one historical example.

(Refer Slide Time: 00:18)



22/36

A HISTORICAL EXAMPLE

- During the Third Reich, the Holocaust was enabled in significant measure by the information technology of the time.
- Hollerith machines, a pre-electronic computer system made and created by Dehomag, the IBM subsidiary in Europe, made it possible for the Nazis to accurately identify Jews using accurately tabulated census data.
- The IBM company under the leadership of Thomas J. Watson actively collaborated with the Nazis in this matter.

NPTEL

iiitb

iiitb  
ज्ञानमुचयते

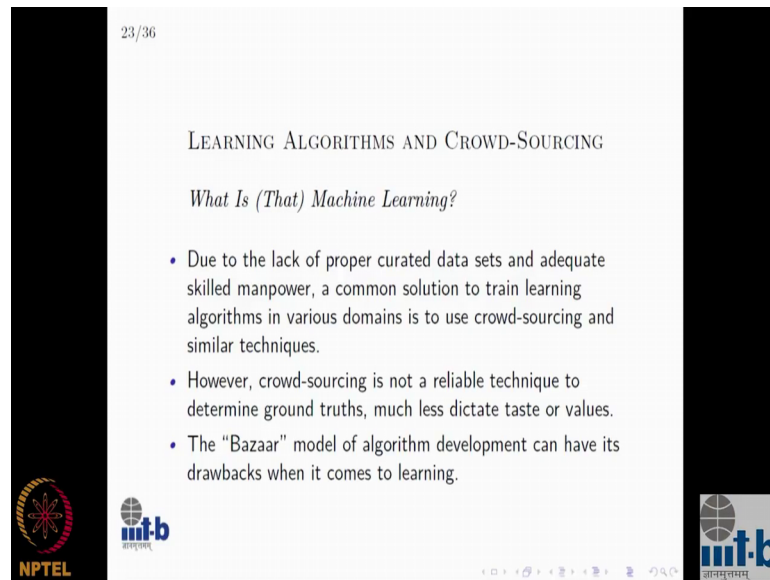
Once again quite well known which is many people do not realize that IT has being around for long time. Did you know the word computer actually referred to a human being first. A computer was a professional word like lawyer or accountant. So, a computer has existed for a very long time in referring to the job done by a human being.

So, you could be a computer just the way you could be an Accountant or a Lawyer or a Doctor. So, back in the day before it electronic it existed there were still hollerith machines and so, on Dehomag the IBM subsidiary in Europe. And there was a lot of senses data collected in Europe especially in Germany during Walmar republic and so on. And IBM a word with Nazis to accurately identify Jews using senses data and that helped the Jews carry out their programs and their holocaust.

And there was a very well known book written about this which is in the references at the end which you can look up. And this is actually a bit of sorry history and IT actually tells you one important lesson to be learned that IT is not always clean IT has not

historically always being clean. And I hope of course, in the entirety of life times and all of human history to come; this sort of things never happen, but it can happen, it has happened once and that something that we need to keep in mind.

(Refer Slide Time: 01:47)



23/36

LEARNING ALGORITHMS AND CROWD-SOURCING

*What Is (That) Machine Learning?*

- Due to the lack of proper curated data sets and adequate skilled manpower, a common solution to train learning algorithms in various domains is to use crowd-sourcing and similar techniques.
- However, crowd-sourcing is not a reliable technique to determine ground truths, much less dictate taste or values.
- The "Bazaar" model of algorithm development can have its drawbacks when it comes to learning.

NPTEL

mitb

mitb

And then machine learning; once again a buzz phrase that many of us have heard and some of us may vertnon also. So, how is machine learning working? In many cases we have we do not really have good curated data set. We do not really know what to learn and how to learn. So, in many cases we are actually using crowd sourcing and so, on we use Amazon, Mechanical Turk and various other ways of doing it.

But crowd sourcing is not a very good way to learn ground rules much less to learn good values and good taste. And there was a very well known example of this just couple of years ago where Microsoft of our company put out a bart which would learn it was a twitter bort; behaviour from other user and became very badly behaved bort, it started using some very foul language and so, on recess language tough like that. So, this is the bazar module there is very well known book written at the; I think late 1990's around 1999 called theatre and the bazaar.

These are two different malls of development especially for IT, but in journal where the cathedral model is there was some fourteenth century French architect who came up with the model who came up with the design for a cathedral and for more than 100 years; people followed that design and built grand cathedral according to that design. The

bazaar model is go in a bazar no real structure there is just pick up pieces where ever you can and then you build something with it using the ideas of different people.

So, the bazaar model of algorithm development also has drawbacks because you are actually giving; giving up a lot of your freedom and you are getting answers from many other people who many or may not have the right answers themselves.

(Refer Slide Time: 03:28)

24/36

LEARNING ALGORITHMS AND  
CROWD-SOURCING—CONT'D

- “The blind application of machine learning runs the risk of amplifying biases present in data.” – Tolga Bolukbasi, *et al.*
- For instance, word embeddings, a technique to find implicit correlations in text data, is widely used in machine learning.
- It makes mappings such as identifying “man” with “computer programmer” and “woman” with “homemaker.”

NPTEL

mit-b

mit-b

And this is once again a very well known example this is actually a code from that blind application machine learning runs the risk of amplifying biases present in data and then this is one very well known paper from nips last year, where they find that there are implicit biases correlations in word embedding. So, for example, a man is identified as a computer programmer and woman is identified as homemaker.

And why is this happening by the way? By the we know that this is wrong, but why is it happening because in fact, most men most computer programmers are men I am not saying that is exactly the way it should be then most women are homemakers. And then the algorithm is actually judging that based on the facts, it is not making value judgment about this at all. But then this word embedding this mapping in some senses is not quite what we wanted to be that is what is happening here.

(Refer Slide Time: 04:22)

25/36

### F/OSS vs. PROPRIETARY SOFTWARE

- In the “Cathedral” model of software development, software is developed by a centralized design process, with a fixed structure.
- Such software is typically proprietary and not open for inspection.
- Free and open-source software (F/OSS) is developed in the “Bazaar” model, often with little centralized control.
- F/OSS source code is open to inspection and code can be inspected, re-used (subject to F/OSS licenses) and forked.

NPTEL

mit-b

mit-b

जलभारतम्

And I just mentioned the bazaar model the other model is of course, cathedral model where software is created by a centralized design process with a certain structure. And almost all of your major operating systems like IOS and even windows and other things they are all built in this.

And such software is typically proprietary we do not have really any way to inspect what is going on this item that may be all kinds of hooks and back doors and questionable ethics and so, on happening inside there and F OSS software free and open sources software is in the bazaar model very little centralized control where you have some code being opened to inspection.

And then we can actually change something, we can actually forth a project and come up with different way of doing that unfortunately all these does not really happen too often. In theory that code is open for inspection, but how many of us are actually honestly looked at Linux kernel for example, have you? I know I will not done that in years and even when I did it was probably not that closely right.

So, the bazaar model is hypothetically better in some cases; it allows you to do more things. it gives you more of your democratic rights and so, on, but typically there are not people investing enough time to make this work very well.

(Refer Slide Time: 05:39)

26/36

### SURVEILLANCE AND PRIVACY

- Algorithms can also be used to track people's movements and behavior (esp. online) much more easily than was formerly possible.
- People's privacy may be invaded dozens of times each day, and "police state" intrusions are quite easy.
- Even more troubling, large corporations have access to huge amounts of private data, with few, if any, restrictions on their use (including reselling to others).

NPTEL IIT Bombay

And then next thing is about surveillance and privacy. So, algorithms can also be used to track peoples movements and behaviour. You know that you must have noticed this many times when you visit website like Flipkart or something and you buy an item you see that you see the Ad for same item in so, many other websites that you visit. So, how is that happening?

Student: (Refer Time: 05:59).

Cookies where you visit a website and they place a cookie and then throughout your web sojourn; your behaviour can be tracked and it notes exactly what you are trying what you are up to. And that is actually in some senses invasion of your privacy it knows your behaviour and by the way that is something you probably do not share even with your close friends; do you always tell your friends where you are visiting all the websites that you visit in order?

Some of that you would say not any of your business what I am doing, but then these online companies actually know that and they track you online so, much. And then you also get photographs many times a day thanks to CCTV and so, on. So, and because all of us carry phones right these days and phones are always on, you can or some company can find out my location accurate within few meters all hours of day and night think about that. Same for all of us even your spouse or your best friend would not know that much about you. So, people's privacy is invaded dozens a times a day and.

Student: (Refer Time: 07:07) we see like we know Google maps and all they collect data from the day whenever my friend location services are on, but if if we do not provide that much data will be will they you know get us correct prediction, correct recommendation if we do not supply them with data; if we know keep our location services off and.

That is certainly the argument that these companies would make then there is some justification for that argument I do not deny. But the counter argument to that is that number one they are often collecting data quite in excess of what they need to provide those services that is not completely true. So, it is it is accurate to say yes you need to give up a little bit of privacy. So, that you can get a service right that part is ok, but you do not always need to give up as much of your privacy as they are demanding to get the service you are getting that is one.

Second thing is that having collected the data why do they then hold it? They do not actually make it available to anyone else, they do not actually release it even to you. Can you ask for your data back? Can you ask can you say I wanted to erase all my data from all your server and get rid of it. And can you say give me all the data about myself I after all I think India; I do not what the status of law is I am not really up to speed on that may be some of you who are studying law you can tell me, but in Europe for example, they are saying that the person individual is the owner of their data. So, they can actually request not just request, they can demand the company erase their data and they can also get their data from the company if they want.

Student:, But the services we are using like the Gmail we use the Whatsapp we use these services are like completely free. So, basically as I say if you are not paying for these services you are actually the product. So, I mean we cannot in a way we cannot ask them to erase the data because they are you know providing us the services for free. So, I mean they can you know like this much amount of data they need consent point, they do they always ask for consent you know.

No they typically what happens is there is a long check list and say agree and do not agree. You always click agree and go on you do not actually bother to read that seven or 8 worth pages text of all and even if you do someone like me certainly does not really understand most of what we read right that is the big problem. And yes I agree there is no

free lunch; if you are actually getting a service then you are; obviously, giving up something for it.

At the same time for example, would a company like Hotmail or Google give us a choice of saying I will pay a 100 dollars a month or whatever it is or may be 100 dollars a month that would be too high 100 rupees a month; let us say with that. And say and do not track me, do not do anything that I do not want you to do; do not deny my privacy at all, but I pay for the service exactly what you are charging, what you are getting from me otherwise. They should snap up that opportunity if they are really honest about it; in practice probably they will not.

Student: They cannot guarantee directly, but the kind of app like you know Gmail if have corporate account.

I do have I in fact many company and in fact our institute has not just Gmail; we have used to have Gmail now we have Office 365. I have a Gmail account, corporate Gmail account through somebody else so, but anyway. So, there are such things, but fact to the matter is this I do not think has a close loop answer at the moment; it is something we need to discuss, it is something we need to think about that is what I am saying.

Student: Coming up with.

Right yes.

Student: No (Refer Time: 10:52) that I mean the take away from are if billions of people are making these simultaneously for the of large volumes of data large scale science like consequences of that and the examples like in terms of in terms of the enforcing (Refer Time: 11:13) large scale governance related issues; what happens when the entity is is not just corporation, but the government we have a we have like a the ethical the question becomes much more apparent.

Student: Really what we getting from you is that there are when simultaneously if we are making individual ethical choices there are large scale (Refer Time: 11:39).

There are issues which are you know; I in fact thank you I think your points are well taken, but my point once again from my perspective is in many cases people like me people who are trained there way I was; I went all the way to a PhD in a computer

science, I studied at visibly good schools without even having to think about this where nobody ever told me that you actually you need to think about how algorithms work in a social context, what the effect of your work or somebody else's work like this will be ah.

Where the consequences of society actually meet people's expectations, where not just technological, but social aspects also play a role. And what I am realizing and what I think that is the realization that I want to share with you is you should think about these things. And that is why it is also of course, quite interdisciplinary it not just computer science issue.

Anyway so coming forward let us I think move on and of course, reselling is a big issue reselling data without telling you a lot of people have objections to that were you have start getting phone calls from companies; that you have never heard of why is that happening? Because somebody who got the data from you, you dropped your call some somebody's fish bowl or something were you shared your number with somebody; then sold you data to somebody else.

Once again think about that you probably do not give out your cell number that easily to just anyone who asks, you do not put it in public places, you do not share it with anyone who ask, but somebody else who has that data is selling it without asking you. Certainly I would find that and I do find that quite objectionable that is not good ok.

(Refer Slide Time: 13:24)

27/36

OTHER ISSUES

*What algorithms can also be used for.*

- Censorship, restrictions on freedom of speech (e.g., the "Great Firewall of China").
- Digital rights management—glorified name for restrictions on sharing.
- Hidden restrictions on democratic rights, meddling with voting systems.

NPTEL

mit-b

mit-b

ज्ञानमुद्रसम्



Then other issues which I will briefly go over. Censorship once again that is very easy to do because if you are only dependent on the internet for your information and you do not actually have physical resources at all; then you can be censored very easily. Because suppose Google says; I will not result I will not show any results that actually deal with a particular political philosophy.

I will not show any work that refers to this school of thought then they can we do not even know about it. And restriction of freedom of speech once again this is the great firewall of China the gentleman in the morning was talking about this were china does not allow Google or Bydo; they do not allow Google or Facebook or any of that they have the firewall were they censor all criticism of the communist party and so on. And that can happen in more certain ways not just in China, but everywhere else also.

Digital rights management that is termed that came up a lot once again around year 2000 because of all these snapstor and online music mp threes the MP3 format was the major breakthrough because it allowed people like you and me to copy a high quality audio file and play music on our personal system; that was why it was a big break through.

And then the recording industry in US was quite alarmed; the recording industry is a profit making industry; they do not they give lot of money to well known recording artist, but otherwise it is not a fare stakeholder. And they actually enforce digital rights management and they prevent innovation because they want to have access to all the content.

They do not necessarily do this out of any greater motive they are not there are suppose as ethical or unethical as most companies are they are driven by corporate greed; they are not they are not trying to do a bad job for themselves. So, therefore, they actually have restrictions on sharing and of course, you can also have restrictions on democratic rights. Once again if you are censored online; your opinions are not allowed to be expressed that is the restriction on democratic rights. And similarly if you have voting systems that use internet or some aspect of the internet and you find a way to medal with them then that also can be a problem.

(Refer Slide Time: 15:38)

28/36

### MECHANISMS AND POLICIES

- It is not commonly appreciated that success in service systems requires good *mechanisms* as well as appropriate *policies*. There is a tendency to focus on mechanisms only.
- The classical example to illustrate the distinction is with locks and keys.
- The technology that allows us to lock doors, cabinets, etc., is an example of a mechanism. The distribution of different keys to various people in an organization, in line with their access privileges, is a policy.

NPTEL MIT-B IIT-B

And now a couple of closing comments this was something that my good friend Vidisha also talked about in the morning. In general in service system you need good mechanism as well as policies and once again computer scientist and so, on tend to focus only on mechanisms. And the classic example of this is locks and keys; now if you have the technology that allows you to lock a door or by any technology. Now that can be the classical locking technology which we have for 100 of years and it can be some swipe card or something else that is a mechanism.

Now, who has which key to which door that is policy should I have the key to your office? Should you have the key to my office? Should we have a key to this building? Should somebody have the key to somebody else have the key to close it; that is policy there is nothing in the mechanism itself that dictates what the policy should be and that is the important distinction to make ah. So, the access privileges that you give that is the policy and you need to make that choice based on the policy and you need to find appropriate policy in certain context.

(Refer Slide Time: 16:44)

29/36

MECHANISMS AND POLICIES—CONT'D

- Not every policy can be enforced given a mechanism, and not every policy that can be implemented given a mechanism is actually appropriate.
- Policies have to evolve as social mores and standards change.
- A mechanism can be agnostic to policy in some ways; for instance, it is possible to segregate resources by gender, age, or handicap status (which is considered quite appropriate) or by race (which would be considered grossly inappropriate).

NPTEL IIT-BOMBAY

And not every policy can be enforced given a mechanism that is important. So, some times you may have a policies that you think very well founded, which are very sound policies, but the technology does not exists for them yet. And once again let me go back to my earlier remark saying that you need the technology and the economic feasibility. So, there may be certain policies which require technology that is not economically feasible. So, for example, I can have a policy that says everyone should completely safe on the road; there should never be any accident; that means, every car should be built like a tank. So, in the event of any collision there will be no hurting anyone inside, but that will be economically infeasible that that car will never actually sell in the market.

So, not every policy can be enforce given a mechanism not every policy that can be implemented is actually appropriate; the other side is also true. So, sometimes mechanisms allow you to implement certain policies which are not the right policies to implement. So, policies of course, also have to evolve as our society evolves and so, in mechanism is also agnostic to policy.

For example, we have segregation of restrooms by gender men's and women's, but you do not want to segregate by race whites only and non whites that used to happen in the US and some other countries not too long ago may be 100 years ago; even more recently. And that we would say completely no; you do not want that you do not tolerate that that is not regarded as appropriate in any way, but that mechanism would exactly the same

both cases. So, policies have an important role to play, you cannot just think about the technology that is used, you also have to think about policy that is being enforced. So, this is my absolute last slide ah.

Algorithm development and software engineering and so, on generally speaking they consider technical problems only. And once again that is the kind of education that I have; in fact, for the most part impart to my own students where we think about the technical aspects of our algorithms work and how certain things need to be constructed, but that is all at the component level.

(Refer Slide Time: 18:50)

30/36

### CONCLUSION

- Algorithm development, and related aspects such as software engineering and the architecture and building of computing services, are generally considered technical problems only.
- It is however essential to appreciate the difference between *mechanisms* (technologies) and *policies*, and let the use of the former be guided by the latter.
- Algorithms and their ilk have the capability to hurt as well as help society and the individual. They must be created and used within a clear and strong ethical framework.
- Laws and policies for the ethical creation and use of algorithms are generally behind the technology, but this has to change.

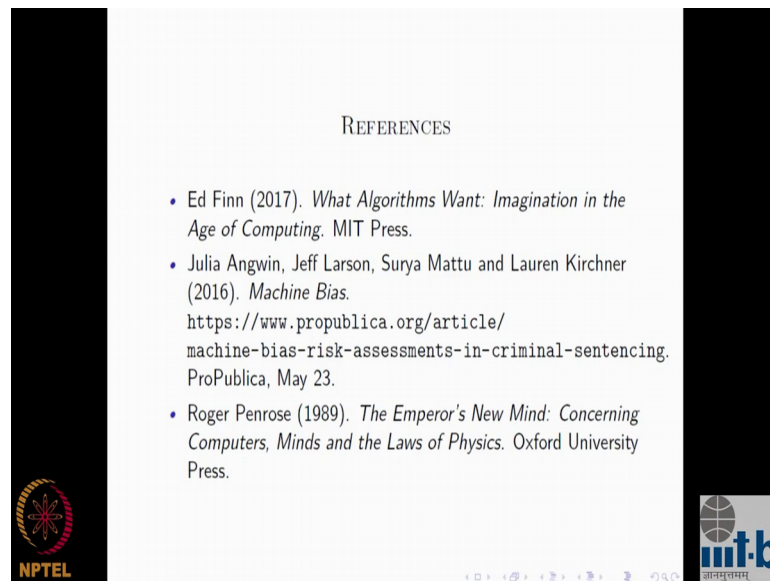
NPTEL IITB

We need to think about that beyond that and think about building computing services not just individual components. And we have to appreciate the difference between mechanisms and policies and let the policy guide how the mechanism is actually deployed. And algorithms have the tendency to help as well as to hurt they can do both, you need to think about that as you go long you cannot just trust the algorithm will always do the right thing.

And laws and policies for all these are actually well behind the curve; technology was very fast laws change very slowly that is also is the that also is the problem I think something that some of you might know more than me. And this has also has to change this is the impact for us to improve society where we have better policies and better laws to do certain things. Any questions? I will be happy to take them any further questions.

I will also briefly go over let me do that I will just mention the references.

(Refer Slide Time: 19:50)



I I suppose some point is slide deck can be made available some references can be made available to them also. This is book that came earlier this year about how algorithms actually have a social context; I do not think this book is actually the word on the subject, but it is a very interesting book; you should try to look at it sometimes if you can.

And this is the very well known pro public article from last year were they actually went into this was actually a big news item at the time, where they found that sentencing for criminal defendants was actually biased when it was done by an algorithms. So, that was this.

And this Roger Penrose book is actually a few decades now Penrose is actually very well known physicist and he is a colleague and friend of Steven Hawking whom I think many of us heard of. And he wrote a book in response to the claims about strong AI where people said that human beings are nothing, but machines and turing machines will eventually do everything that person can and all that stuff.

So, then you heard of the famous story of the Emperors new cloths? Yes the emperor who was supposed to be wearing this really fine cloths turned out to be not to be not wearing anything. So, that was the pun that he is made here the Emperors new mind. So,

he is saying that you are saying that in AI system actually has a mind, but that is the Emperors new mind; there is no real mind there that is what this is that is the pun in the title.

(Refer Slide Time: 21:17)

REFERENCES—CONT'D

- Ted Striphas (2015). *Algorithmic Culture*. European Journal of Cultural Studies, Vol 18, Issue 4–5, pp. 395–412. <http://journals.sagepub.com/doi/abs/10.1177/1367549415577392>
- Edwin Black (2001). *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation*. Crown Books.
- Utpal M. Dholakia (2015). *The Perils of Algorithm-Based Marketing*. Harvard Business Review, June 2015. <https://hbr.org/2015/06/the-perils-of-algorithm-based-marketing>

The slide features the NPTEL logo on the bottom left and the IIT Bombay logo on the bottom right. A navigation bar with various icons is visible at the bottom center.

And then this is known this is risibly good paper actually he is actually only gives you bunch of definitions, but it is still you can look at this for some inside it what algorithmic culture is. And this is a very a well known book from many years ago not too many, but earlier this century where he came up with the study on the how IBM actually collaborated with the Nazis on identifying Jews and things like that.




And there are some there are numerous things about algorithm based marketing and how they are not very good and this is one example of that this is by paper

(Refer Slide Time: 21:54)

34/36

REFERENCES—CONT'D

- Daniel Victor (2016). *Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk*. New York Times, March 24, 2016. <http://tiny.cc/nytimes-ms-bot>.
- Eric Raymond (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media.
- Jeremy Tammik (2016). *Richard Stallman for a Free Digital Society*. Blog post, February 2016. <http://tiny.cc/stallman-talk>.



And this is the news article in the New York Times about how Microsoft twitter brought actually misbehaved once they let it go they actually had to pull it down after about a day because it learned behaviour from many bad people; people who are misbehaving online and Eric Raymond he is he is open source vocationery and he actually advocates for this and he came up with this book.

And Richard Stallman of course, many of us know who Richard Stallman is; he is one of the gurus of open source also and he has these things called free software foundation and Gnu. So, actually he has a video here that is the link for that.

(Refer Slide Time: 22:36)

35/36

REFERENCES—CONT'D

- Gregory R. McArthur (1985). *If Writers Can't Program and Programmers Can't Write, Who's Writing User Documentation?* SIGDOC '85: Proceedings of the 4th Annual International Conference on Systems Documentation, Ithaca, NY, pp. 62-70.
- John Simon (1980). *Compact With Computers*. In *Paradigms Lost: Reflections on Literacy and Its Decline*. Random House.
- Michael Sipser (2005). *Introduction to the Theory of Computation* (2d ed.). Course Technology. ISBN 978-0534950972.

NPTEL

MIT-B

MIT-B

And that is the very well known title from the 1980s *If Writers Cannot Program and Programmers Cannot write, Who is Writing User Documentation?* And then in the 1980's even before there was actually this is actually a very interesting historical article. Because it turns out that there was very serious claims about how would AI would be; even back in the 1970's and 1980's and in the 1970's someone actually wrote saying that in the future, you will never need to be educated because computers will do all the thinking for you.

And they were saying about the year 2000 in the year 2000 which is now well in our past, they were saying you will not need to know how to read because the machine will do the reading for you. So, anyway this guy in 1980, he wrote an article which critic that and its very funny and very well written article.

And Sipser is book if you I think many of us who have done some CS; you know this already it is a book that you that you look at about theoretical machine stuff like that.



(Refer Slide Time: 23:31)

36/36

REFERENCES—CONT'D

- Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai, 2016. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. NIPS 2016. arXiv:1607.06520 [cs.CL].
- US District Court Judgment, case 4:14 cv 01189, May 2017. <https://www.courthousenews.com/wp-content/uploads/2017/05/HoustonTeachers.pdf>
- Mark Bauerlein, 2011. *The Digital Divide: Arguments for and Against Facebook, Google, Texting, and the Age of Social Networking*. TarcherPerigee.

NPTEL

MIT-B

MIT-B

And this is the famous NIPS paper; man is to computer programmer as woman is to homemaker; it caused big splash last year when it came out last year. And this is the district court judgement about Houston schools district that I mentioned in the beginning. And this is an article about digital device not an article; it is a book actually about digital device anyway thank you and any questions feel free.

Student: (Refer Time: 24:02).

I do not think there is a no speaker here they are using that.

Student: (Refer Time: 24:07).

That is only for the recording yes even my mike does not actually go to any speaker as far as all.

Student: Sir I was interested in the crowd sourcing part that you are talking about; the crowd sourcing part the decisions are because like.

Right.

Student: Taken together from crowd sourcing platforms.

Right.

Student: So, regarding that two kinds of decisions one kind of decisions is where people will know the discussion decisions by discussing them proper discussion decision by discussing them has in they have inherent solution which can be discussed.

Right.

Student: And some problems like the autonomous cars and things like that and whom to save and whom to not save very tricky and even after discussion we might not end up with the correct answer or ethical answer that agree everybody agrees on. So, I of thinking crowd sourcing into these two different context like in one context crowd sourcing as I is I might agree that it might not be good. But in the other context where there are autonomous cars and things like that; is it a what I mean I am asking for your opinion if its better way to do it? Because I have I just came across one particular project from MIT it is called mourning machine, where they are trying to build a take open source data from decision from people about training autonomous cars such situations. So, what would you think?

I do not know that much about autonomous cars to be quite honest and to my knowledge to my limited knowledge autonomous cars are not a major use case or domains for crowd sourcing. There may be particular problem there which requires crowd sourcing, but crowd sourcing is used in couple of different cases for example, one case is where they have you know what captchas are? Captchas are those irritating figures that they show us and that is supposed to distinguish between people and robots right.

So, in some cases if you want to prevent the bot from signing up for something and using an account and so, on; they make sure supposedly that you are a person by asking you to solve that captcha. And in fact, now that captcha framework itself is gradually collapsing because deep learning is making it possible to get bots that would actually solve the captcha as well as human being.

But classically at least for the last 15 years captchas have being used for that purpose. So, what Google has done I think the Google books project or some other cases says they actually show you snippets of text. And they ask people to identify what that text reads and they use crowd sourcing to essentially do OCR; not really OCR they get the ground truth for that. So, they will get they will show same image to 10 people and they will get majority opinion about what that text is they will use that in their to get the data for their

OCR or whatever it is. And that that is actually a that is actually a case where crowd sourcing makes sense.

Where it makes less sense is where you have cultural norms or truths or norms of behaviour, where once again like the Microsoft Twitter bot where people who are online may or may not be behaving in the way you want to train your machine to behave. So, if you are learning certain truths like that, if you are learning ground truths or behaviour modes of behaviour that are socially constructed and they are not scientific truth in such cases; you are probably on shaky ground using crowd sourcing.

Using mechanical Turk or something like it for OCR and. So, on or Google for OCR using captcha and so, on there you are more solid footing because those are not cultural constructions, those are either that letter is yell or its not we can agree or disagree if 8 out of 10 people say it is a yell probably a yell we can go with that that part will work, but if a certain word is not appropriate used in polite company, but that bot learns that from a lot of people who are loosely using it they are may be not the right company for that bot at that moment.

Just like a child right you do not want the child to learn foul language therefore, I think those of us who are you guys may be too young for that, but some of us who have children we think about this all the time; I do not want to let loose with lot of foul language in front of small children because they learn from me and they might learn the wrong thing from me.

So, same thing about crowd sourcing when the context of behaviours so, my very rough answer to you would be when you are on solid ground in terms of ground truth; if that is not a social or cultural construction if it is that; it is that clear like letter of text which is in some grainy image that you are trying to decipher that I think crowd sourcing will work very well for it. If it is cultural I would be a little more careful because I do not know the culture of the people from which it is imbibing that.

Student: (Refer Time: 29:04) what has getting it actually. So, we talked about two kinds of ground truth that is one is what we know for sure almost most likely the what it is going to be. And the other one is very shaky based on the cultural norms. So, what I am going with the context no ground truths like trolley problem. So, that in that cases should we rely on crowd information what would people do in such cases?

I suppose, but if you want to stick with autonomous cars and so, on; if you learn ground truth in Bangalore or if you learn the crowd source behaviour in Bangalore and you apply that in New York, you probably end up with some very undesirable conclusions and also the opposite. In the sense that the behaviour or the norms in a particular society including driving behaviour and the behaviour of pedestrians and drivers and so, on can be quite different based on the local society.

Student: So, these (Refer Time: 29:59) can be given a ground truth right, but I am talking about more ethical issues kind of things so.

No, I do not know how you can actually I am not an ethicist and I do not really have good foundation in ethics; may be someone who does can open on this, but I do not think ethics in general are determined based on crowd sourcing because that can become problematic.

Because just like scientific truth are not fixed based on democratic shall I say outcomes right. You cannot actually pole a number of people and find out what a scientific fact is; I do not think and this is me speaking without having the background in ethics that you cannot fix an ethical outcome by a crowd sourced answer. I think it at most tell you what most people think the ethic ethics of the situation is, but that may or may not be the correct answer. You need an ethical frame work I think there are people who studied that; unfortunately I am not one of them. So, I do not have a very good answer for you.

Student: (Refer Time: 31:04) Google you know used captchas do you know create I mean OCR the Google books they had.

Student: So, how did it they validate the text that we entered for the captcha? Because it it needs to you know say yes or no was the captcha correct or not. So, how did it validate like if you entered the captcha was it correct or not because it is you know it is learning from you.

Right.

Student: So, how will it say I mean tell you that you are a human being or not I mean the purpose behind it.

I do not really completely know that because I am not involved in this project and not really studied it, but I know this particular concept has being used in this.

Student: Even I read about it this is the (Refer Time: 31:51) I mean.

I do not know I have look it up because this was I think 4 or 5 years ago.

It was not even very recent it has being done that is what I can tell you, I really do not know the details of how it has happened; I think that lady wants the mike.

Student: I I just a quick comment which is that it is just interesting that if we go back right to the beginning of your talk and like what you began with which is that by definition or by nature itself algorithms are about supplanting human decision making like.

Student: that is I mean the point is to automaticise the certain sequence.

Automate.

Student: Certain sequence of procedures and so, on right. So, it is just now and then if you skip back to end and you are looking at like the problematic; the outcomes much more political social plain that the question just become lot more trickier. Because it almost like by nature the technology is supposed to displace human decisions making. And that it is embodying judgement that; that one does not entirely have control over. And now we are sort of taking issue with that right; so, that push that push and pull between the nature and by definition; the character of the technology itself and then the almost like its logical outcome or.

Right.

Student: The logical consequences I think it has become its just it just very powerful example to show the technology is not neutral and the it is not.

Yes.

Student: That it is not kind of an objective kind of absolute self evident truth, but it is by from the get going embedded in our social.

(Refer Time: 33:43) yes.

Student: Cultural context.

Yes very much so, and sort of is almost a closing thought I would say as I said before one of the big reasons for AI and automation is not; it is not the company companies want to make a profit and that profit often arises when you reduce the work force, when you have a linear work force; yes there is a there is that aspect to it.

But the major push for AI and automation is not only that it is because the once again Facebook is now serving a billion plus people, but their workforce cannot really scale in proportion. So, if you have a company like Facebook that serves only a million people; it will still have probably something close to the same workforce and Facebook cannot really hire 1000 times as many people to serve a much larger workforce, much larger customer base that is the problem. So, in many cases you have to you have to automate to serve the number of customers who actually want the services; you want you are trying to offer and; that means, that there is no really getting away from complex systems and AI.

That is the major push for this and in many cases people who want what you and I would call as intelligent services. We do not really have a very thorough definition for AI, but one definition that it seems to work in many context is; if you would associate a human being with intelligence if he or she did that and if a system does that, if a machine does that then you can all that is being AI; Artificial Intelligence.

So, what do you define as intelligent for a human being? Play chess; plays very good chess, plays or plays music, creates music something else like that something which you think is a skill of or a talent or high quality intelligence. So, if something like that can be done by a system including decision making for some of these context and so, on if such process can be replicated by an algorithm; then that algorithm also has AI that is the or the system has AI to be more precise that is one working definition.

So, we need AI precisely because there are not enough people available to do all those things. There are not enough people for 6, 7 billion of us to be served with all the things we want; that is the catch ok. Thank you being a very fine audience.

Thank you.